

l_0 Norm Constraint LMS Algorithm for Sparse System Identification

Yuantao Gu*, Jian Jin, and Shunliang Mei

Received Mar. 24, 2009; accepted June 5, 2009.

This article appears in *IEEE Signal Processing Letters*, 16(9):774-777, 2009.

Abstract

In order to improve the performance of Least Mean Square (LMS) based system identification of sparse systems, a new adaptive algorithm is proposed which utilizes the sparsity property of such systems. A general approximating approach on l_0 norm – a typical metric of system sparsity, is proposed and integrated into the cost function of the LMS algorithm. This integration is equivalent to add a zero attractor in the iterations, by which the convergence rate of small coefficients, that dominate the sparse system, can be effectively improved. Moreover, using partial updating method, the computational complexity is reduced. The simulations demonstrate that the proposed algorithm can effectively improve the performance of LMS-based identification algorithms on sparse system.

Keywords: adaptive filter, sparsity, l_0 norm, Least Mean Square (LMS).

1 Introduction

A sparse system is defined whose impulse response contains many near-zero coefficients and few large ones. Sparse systems, which exist in many applications, such as Digital TV transmission channels [1] and the echo paths [2], can be further divided to general sparse systems (Fig. 1. a) and clustering sparse systems (Fig. 1. b, ITU-T G.168). A clustering sparse system consists of one or more clusters, wherein a cluster is defined as a gathering of large coefficients. For example, the acoustic echo path is a typical single clustering sparse system, while the echo path of satellite links is a multi-clustering system which includes several clusters.

There are many adaptive algorithms for system identification, such as Least Mean Squares (LMS) and Recursive Least Squares (RLS) [3]. However, these algorithms have

*This work was partially supported by National Natural Science Foundation of China (NSFC 60872087 and NSFC U0835003). The authors are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. The correspond author of this paper is Yuantao Gu (e-mail: gyt@tsinghua.edu.cn).

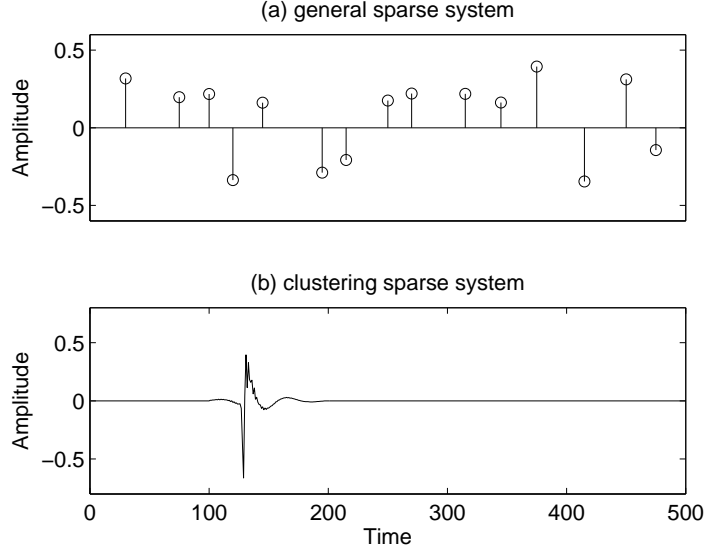


Figure 1: Typical sparse system.

no particular advantage in sparse system identification due to no use of sparse characteristic. In recent decades, some algorithms have exploited the sparse nature of a system to improve the identification performance [2, 4–10]. As far as we know, the first of them is Adaptive Delay Filters (ADF) [4], which locates and adapts each selected tap-weight according to its importance. Then, the concept of proportionate updating was originally introduced for echo cancellation application by Duttweiler [2]. The underlying principle of Proportionate Normalized LMS (PNLMS) is to adapt each coefficient with an adaptation gain proportional to its own magnitude. Based on PNLMS, there exists many improved PNLMS algorithms, such as IPNLMS [5] and IIPNLMS [6]. Besides the above mentioned algorithms, there are various improved LMS algorithms on clustering sparse system [7, 8, 10]. These algorithms locate and track non-zero coefficients by dynamically adjusting the length of the filter. The convergence behaviors of these algorithms depend on the span of clusters (the length from the first non-zero coefficient to the last one in an impulse response). When the span is long and close to the maximum length of the filter or the system has multiple clusters, these algorithms have no advantage compared to the traditional algorithms.

Motivated by Least Absolutely Shrinkage and Selection Operator (LASSO) [11] and the recent research on Compressive Sensing (CS) [12], a new LMS algorithm with l_0 norm constraint is proposed in order to accelerate the sparse system identification. Specifically, by exerting the constraint to the standard LMS cost function, the solution will be sparse and the gradient descent recursion will accelerate the convergence of near-zero coefficients in the sparse system. Furthermore, using partial updating method, the additional computational complexity caused by l_0 norm constraint is far reduced. Simulations show that the new algorithm performs well for the sparse system identification.

2 New LMS Algorithm

The estimation error of the adaptive filter output with respect to the desired signal $d(n)$ is

$$e(n) = d(n) - \mathbf{x}^T(n)\mathbf{w}(n), \quad (1)$$

where $\mathbf{w}(n) = [w_0(n), w_1(n), \dots, w_{L-1}(n)]^T$ and $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-L+1)]^T$ denote the coefficient vector and input vector, respectively, n is the time instant, and L is the filter length. In traditional LMS the cost function is defined as squared error $\xi(n) = |e(n)|^2$.

By minimizing the cost function, the filter coefficients are updated iteratively,

$$w_i(n+1) = w_i(n) + \mu e(n)x(n-i), \quad \forall 0 \leq i < L, \quad (2)$$

where μ is the step-size of adaptation.

The research on CS shows that sparsity can be best represented by l_0 norm, in which constraint the sparsest solution is acquired. This suggests that a l_0 norm penalty on the filter coefficients can be incorporated to the cost function when the unknown parameters are sparse. The new cost function is defined as

$$\xi(n) = |e(n)|^2 + \gamma \|\mathbf{w}(n)\|_0, \quad (3)$$

where $\|\cdot\|_0$ denotes l_0 norm that counts the number of non-zero entries in $\mathbf{w}(n)$, and $\gamma > 0$ is a factor to balance the new penalty and the estimation error. Considering that l_0 norm minimization is a Non-Polynomial (NP) hard problem, l_0 norm is generally approximated by a continuous function. A popular approximation [13] is

$$\|\mathbf{w}(n)\|_0 \approx \sum_{i=0}^{L-1} \left(1 - e^{-\beta |w_i(n)|}\right), \quad (4)$$

where the two sides are strictly equal when the parameter β approaches infinity. According to (4), the proposed cost function can be rewritten as

$$\xi(n) = |e(n)|^2 + \gamma \sum_{i=0}^{L-1} \left(1 - e^{-\beta |w_i(n)|}\right). \quad (5)$$

By minimizing (5), the new gradient descent recursion of filter coefficients is

$$w_i(n+1) = w_i(n) + \mu e(n)x(n-i) - \kappa \beta \text{sgn}(w_i(n))e^{-\beta |w_i(n)|}, \quad \forall 0 \leq i < L, \quad (6)$$

where $\kappa = \mu\gamma$ and $\text{sgn}(\cdot)$ is a component-wise sign function defined as

$$\text{sgn}(x) = \begin{cases} \frac{x}{|x|} & x \neq 0; \\ 0 & \text{elsewhere.} \end{cases} \quad (7)$$

Table 1: The Pseudo-codes of l_0 -LMS

Given L, Q, μ, β, κ ;
Initial $w = \text{zeros}(L,1)$, $f = \text{zeros}(L,1)$;
For $i = 1, 2, \dots$
 input new x and d ;
 $e = d - x' * w$;
 $t = \text{mod}(i, Q)$;
 $f(t+1:Q:L) = -\beta * \max(0, 1 - \beta * \text{abs}(w(t+1:Q:L))) * \text{sign}(w(t+1:Q:L))$;
 $w = w + \mu * e * x + \kappa * f$;
End

To reduce the computational complexity of (6), especially that caused by the last term, the first order Taylor series expansions of exponential functions is taken into consideration,

$$e^{-\beta|x|} \approx \begin{cases} 1 - \beta|x| & |x| \leq \frac{1}{\beta}; \\ 0 & \text{elsewhere.} \end{cases} \quad (8)$$

It is to be noted that the approximation of (8) is bounded to be positive because the exponential function is larger than zero. Thus equation (6) can be approximated as

$$w_i(n+1) = w_i(n) + \mu e(n)x(n-i) + \kappa f_\beta(w_i(n)) \quad \forall 0 \leq i < L, \quad (9)$$

where

$$f_\beta(x) = \begin{cases} \beta^2 x + \beta & -\frac{1}{\beta} \leq x < 0; \\ \beta^2 x - \beta & 0 < x \leq \frac{1}{\beta}; \\ 0 & \text{elsewhere.} \end{cases} \quad (10)$$

The algorithm described by (9) is denoted as l_0 -LMS. Its implementation costs more than traditional LMS due to the last term in the right side of (9). It is necessary, therefore, to reduce the computational complexity further. Because the value of the last term does not change significantly during the adaptation, the idea of partial updating [14] [15] can be used. Here the simplest method of sequential LMS is adopted. That is, at each iteration, one in Q coefficients (where Q is a given integer in advance) is updated with the latest $f_\beta(w_i(n))$, while those calculated in the previous iterations are used for the other coefficients. Thus, the excessive computational complexity of the last term is one in Q th of the original method. More detailed discussion on partial update can be found in [14]. The final algorithm is described using MATLAB like pseudo-codes in TABLE 1.

In addition, the proposed l_0 norm constraint can be readily adopted to improve most LMS variants, e.g. NLMS [3], which may be more attractive than LMS because of its

robustness. The new recursion of l_0 -NLMS is

$$w_i(n+1) = w_i(n) + \mu \frac{e(n)x(n-i)}{\delta + \mathbf{x}^T(n)\mathbf{x}(n)} + \kappa f_\beta(w_i(n)), \quad \forall 0 \leq i < L, \quad (11)$$

where $\delta > 0$ is the regularization parameter.

3 Brief Discussion

The recursion of filter coefficients in the traditional LMS can be expressed as

$$\mathbf{w}_{\text{new}} = \mathbf{w}_{\text{prev}} + \text{gradient correction}, \quad (12)$$

where the filter coefficients are updated along the negative gradient direction. Equation (9) can be presented in the similar way,

$$\mathbf{w}_{\text{new}} = \mathbf{w}_{\text{prev}} + \text{gradient correction} + \text{zero attraction}, \quad (13)$$

where zero attraction means the last term in (9), $\kappa f_\beta(w_i(n))$, which imposes an attraction to zero on small coefficients. Particularly, referring to Fig. 2, after each iteration, a filter weight will decrease a little when it is positive, or increase a little when it is negative. Therefore, it seems that in \mathbb{R}^L space of tap coefficients, *an attractor*, which attracts the non-zero vectors, exists at the coordinate origin. The range of attraction depends on the parameter, β .

The function of *zero attractor* leads to the performance improvement of l_0 -LMS in sparse system identification. To be specific, in the process of adaptation, a tap coefficient closer to zero indicates a higher possibility of being zero itself in the impulse response. As shown in Fig. 2, when a coefficient is within a neighborhood of zero, $(-1/\beta, 1/\beta)$, the closer it is to zero, the greater the attraction intensity is. When a coefficient is out of the range, no additional attraction is exerted. Thus, the convergence rate of those near-zero coefficients will be raised. In conclusion, the acceleration of convergence of near-zero coefficients will improve the performance of sparse system identification since those coefficients are in the majority.

According to the above analysis, it can be readily accepted that β and κ determine the performance of the proposed algorithm. Here a brief discussion about the choice of these two parameters will be given.

- The choice of β : As mentioned above, strong attraction intensity or a wide attraction range, which means the tap coefficients are attracted more, will accelerate the convergence. According to Fig. 2, a large β means strong intensity but a narrow attraction range. Therefore, it is difficult to evaluate the impact of β on the convergence rate. For practical purposes, Bradley and Mangasarian in [13] suggest to set the value of β to some finite value like 5 or increased slowly throughout the iteration process for better approximation. Here, $\beta = 5$ is also proper. Further details are omitted here for brevity. And readers of interest please refer to [13].

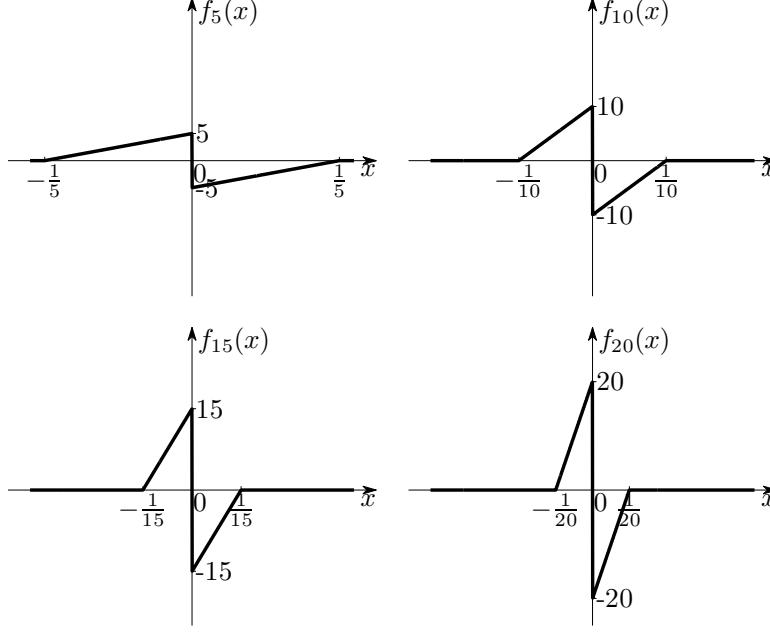


Figure 2: The curves of function $f_\beta(x)$ with $\beta=5, 10, 15, 20$.

- The choice of κ : According to (3) or (9), the parameter κ denotes the importance of l_0 norm or the intensity of attraction. So a large κ results in a faster convergence since the intensity of attraction increases as κ increases. On the other hand, steady-state misalignment also increases as κ increases. After the adaptation reaches steady state, most filter weights are near to zero due to the sparsity. We have $|\kappa f_\beta(w_i(n))| \approx \kappa\beta$ for most i . Regarding to those near-zero coefficients, $w_i(n)$ will move randomly in the small neighborhood of zero, driven by the attraction term as well as the gradient noise term. Therefore, a large κ results in a large steady-state misalignment. In conclusion, the parameters κ are determined by the trade-off between adaptation speed and adaptation quality in particular applications.

4 Simulations

The proposed l_0 -NLMS is compared with the conventional algorithms NLMS, Stochastic Taps NLMS (STNLMS) [7], IPNLMS, and IIPNLMS in the application of sparse system identification. The effect of parameters of l_0 -LMS is also tested in various scenarios. $\beta = 5$ and the proposed partial updating method with $Q = 4$ for l_0 -LMS and l_0 -NLMS is used in all the simulations.

The first experiment is to test the convergence and tracking performance of the proposed algorithm driven by a colored signal. The unknown system is a network echo path, which is initialized with the echo path model 5 in ITU-T recommendation, delayed by 100 taps and tailed zeros (clustering sparse system, Fig. 1. b). After 3×10^4 iterations, the delay

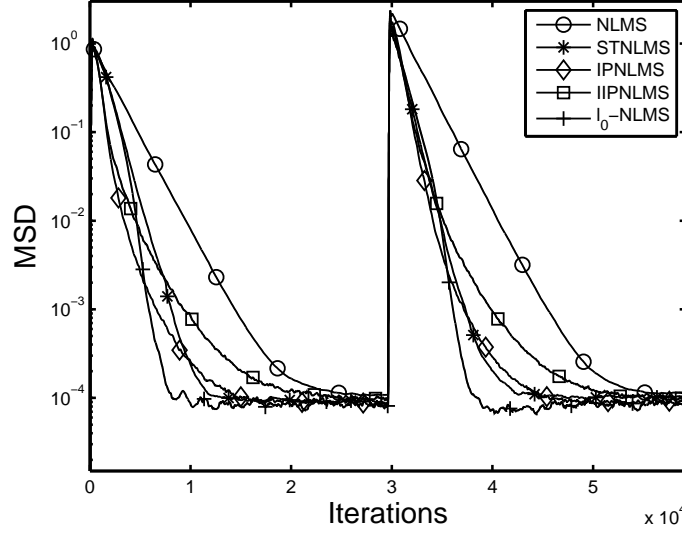


Figure 3: Comparison of convergence rate for five different algorithms, driven by colored signal.

is enlarged to 300 taps and the amplitude decrease 6dB. The input signal is generated by white Gaussian noise $u(n)$ driving a first-order Auto-Regressive (AR) filter, $x(n) = 0.8x(n-1) + u(n)$, and $x(n)$ is normalized. The observed noise is white Gaussian with variance 10^{-3} . The five algorithms are simulated for a hundred times, respectively, with parameters $L = 500, \mu = 1$. The other parameters as follows.

- IPNLMS [5] and IIPNLMS [6]: $\rho = 10^{-2}, \alpha = 0, \alpha_1 = -0.5, \alpha_2 = 0.5, \Gamma = 0.1$;
- ST-NLMS: the initial positions of the first and last active taps of the primary filter are 0 and 499, respectively; those of the auxiliary filter are randomly chosen;
- l_0 -NLMS: $\kappa = 8 \times 10^{-6}$;

Please notice that the parameters of all algorithms are chosen to make their steady-state errors the same. The Mean Square Deviations (MSDs) between the coefficients of the adaptive filter and the unknown system are shown in Fig. 3. According to Fig. 3, the proposed l_0 -NLMS reaches steady-state first among all algorithms. In addition, when the unknown system abruptly changes, again the proposed algorithm reaches steady-state first.

The second experiment is to test the convergence performance of l_0 -LMS with different parameters κ . Suppose that the unknown system has 128 coefficients, in which eight of them are non-zero ones (their locations and values are randomly selected). The driven signal and observed noise are white, Gaussian with variance 1 and 10^{-4} , respectively. The filter length is $L = 128$. The step-size of l_0 -LMS is fixed to 10^{-2} , while κ is with different values. After a hundred times run, their MSDs are shown in Fig. 4, in which MSDs of LMS

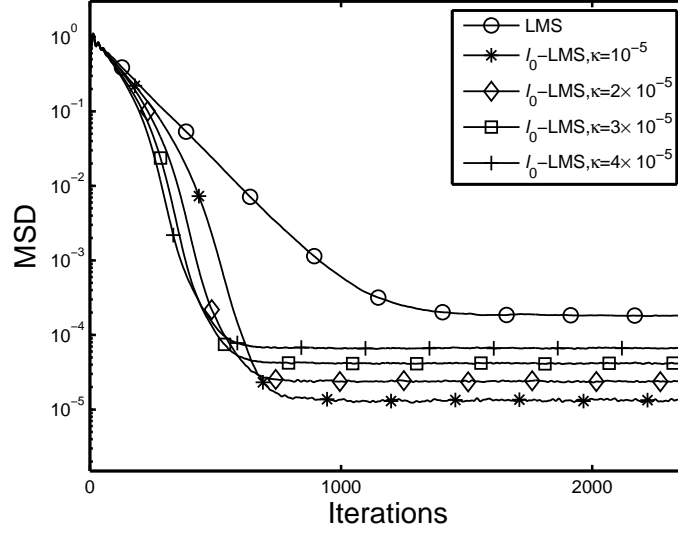


Figure 4: Learning curves of l_0 -LMS with different κ , driven by white signal.

($\mu = 10^{-2}$) are also plotted for reference. It is evidently recognized that l_0 norm constraint algorithm converges faster than its ancestor. In addition, from the figure, it is obvious that a larger κ results in a higher convergence rate but a larger steady-state misalignment. These illustrate again that a delicate compromise should be made between the convergence rate and steady-state misalignment for the choice of κ in practice. Certainly, the above results are consistent with the discussion in the previous section.

The third experiment is to test the performance of l_0 -LMS algorithm with various sparsities. The unknown system is supposed to have a total of 128 coefficients and is a general sparse system. The number of large coefficients varies from 8 to 128, while the other coefficients are Gaussian noise with a variance of 10^{-4} . The input driven signal and observed noise are the same as that in the first experiment. The filter length is also $L = 128$. In order to compare the convergence rate in all scenarios, the step-sizes are fixed to 6×10^{-3} . Parameter κ is carefully chosen to make their steady-state error the same (TABLE 2). All algorithms are simulated 100 times respectively and their MSDs are shown in Fig. 5. As predicted, the number of the large coefficients has no influence on the performance of LMS. However, the convergence rate decreases as the number of large coefficients increases for l_0 -LMS. Therefore, the new algorithm is sensitive to the sparsity of system, that is, a sparser system has better performance. As the number of large coefficients increases, the performance of l_0 -LMS is gradually degraded to that of standard LMS. Meanwhile, it is to be emphasized that in all cases, the l_0 -LMS algorithm is never worse than LMS.

Table 2: The parameters of l_0 -LMS in the 3rd experiment.

LCN ¹	8	16	32	64	128
κ	8×10^{-5}	5.5×10^{-5}	4.5×10^{-5}	3.5×10^{-5}	10^{-6}

[1]LCN denotes Large Coefficients Number.

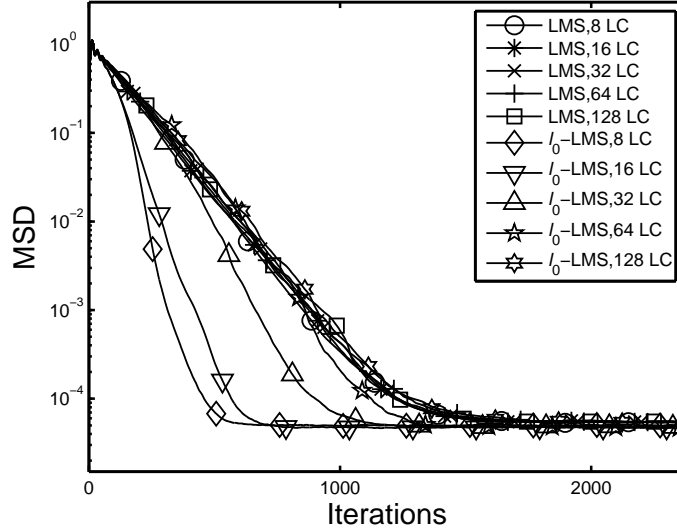


Figure 5: Learning curves of l_0 -LMS and LMS with different sparsities, driven by white signal, where LC denotes Large Coefficients.

5 Conclusion

In order to improve the performance of sparse system identification, a new LMS algorithm is proposed in this letter by introducing l_0 norm, which has vital impact on sparsity, to the cost function as an additional constraint. Such improvement can evidently accelerate the convergence of near-zero coefficients in the impulse response of a sparse system. To reduce the computing complexity, a method of partial updating coefficients is adopted. Finally, simulations demonstrate that l_0 -LMS accelerates the identification of sparse systems. The effects of algorithm parameters and unknown system sparsity are also verified in the experiments.

Acknowledgment

The authors are very grateful to the anonymous reviewers for their part in improving the quality of this paper.

References

- [1] W. F. Schreiber, "Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions", *Proc. IEEE*, vol. 83, no. 6, 1995, pp. 958-981.
- [2] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers", *IEEE Trans. on Speech Audio Process*, vol. 8, no. 5, 2000, pp. 508-518.
- [3] B. Widrow and S. D. Stearns, *Adaptive signal processing*, New Jersey. Prentice Hall, 1985.
- [4] D. M. Etter, "Identification of sparse impulse response systems using an adaptive delay filter", *ICASSP 85*, pp. 1167-1172.
- [5] J. Benesty and S. L. Gay, "An improved PNLMS algorithm", *Proc. IEEE ICASSP*, 2002, pp. II-1881-II-1884.
- [6] P. A. Naylor, J. Cui, M. Brookes, "Adaptive algorithms for sparse echo cancellation", *Signal Processing*, vol. 86, no. 6, pp. 1182-1192, 2006.
- [7] Y. Li, Y. Gu, and K. Tang, "Parallel NLMS filters with stochastic active taps and step-sizes for sparse system identification", *ICASSP 06*, vol. 3, pp. 109-112.
- [8] V. H. Nascimento, "Improving the initial convergence of adaptive filters: variable-length LMS algorithms", *DSP 2002*, Vol. 2, pp. 667-670
- [9] R. K. Martin, W. A. Sethares, et al., "Exploiting sparsity in adaptive filters", *IEEE Trans. Signal Processing*, vol. 50, pp. 1883-1894, Aug. 2002.
- [10] O. A. Noskoski and J. Bermudez, "Wavelet-Packet-Based Adaptive Algorithm for Sparse Impulse Response Identification", *ICASSP*, vol. 3, pp. 1321-1324, 2007.
- [11] R. Tibshirani, "Regression shrinkage and selection via the LASSO". *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 267-288, 1996.
- [12] D. Donoho, "Compressed sensing", *IEEE Trans. Inform. Theory*, vol. 52, no. 4, Apr. 2006, pp. 1289-1306.
- [13] P. S. Bradley and O. L. Mangasarian, "Feature selection via concave minimization and support vector machines", *Proc. 13th ICML*, 1998, pp. 82-90.
- [14] C. D. Scott, "Adaptive Filters Employing Partial Updates", *IEEE Trans. Circuit and Systems*, vol. 44, no. 3, pp. 209-216, 1997.
- [15] M. Godavarti and A. O. Hero, "Partial update LMS algorithms", *IEEE Trans. on Signal Processing*, vol. 53, no. 7, 2005, pp. 2382-2399.